

## 4FX: félig kompozicionális szerkezetek automatikus azonosítása többnyelvű korpuszon

Rácz Anita<sup>1</sup>, Nagy T. István<sup>1</sup>, Vincze Veronika<sup>2</sup>

<sup>1</sup>Szegedi Tudományegyetem, TTK, Informatikai Tanszékcsoport,  
Szeged Árpád tér 2., e-mail: raczanita89@gmail.com, nistvan@inf.u-szeged.hu

<sup>2</sup>Magyar Tudományos Akadémia, Mesterséges Intelligencia Kutatócsoport,  
Szeged, Tisza Lajos körút 103., e-mail: vinczev@inf.u-szeged.hu

**Kivonat** Jelen tanulmányunkban ismertetjük gépi tanulási módszeren alapuló megközelítésünket, mely segítségével négy nyelvű párhuzamos korpuszon automatikusan azonosítottuk a félig kompozicionális szerkezeteket (FX). Elsőként felderítettük a lehetséges jelölteket a magyar, angol, német és spanyol jogi szövegekben, majd egy gazdag jellemzőkészleten alapuló bináris osztályozó segítségével azonosítottuk e szerkezeteket. Ennek az alapvetően adatvezérelt módszernek az alapja a manuálisan annotált 4FX korpusz. Ezáltal lehetőségünk nyílik az FX-ek nyelvspecifikus sajátosságainak vizsgálatára. A 4FX korpusz, illetve a négy nyelvre megvalósított automatikus detektáló hozzájárulhat számos számítógépes nyelvészeti alkalmazás, például gépi fordítók hatékonyságának javításához is.

**Kulcsszavak:** információkinyerés, természetesnyelv-feldolgozás, felszíni szintaktikai elemzés

### 1. Bevezetés

A félig kompozicionális szerkezetek (FX) az összetett kifejezések egyik típusa, melyek egy igei és egy főnévi komponensből épülnek fel. A főnév főként a szemantikai funkciókért, míg az ige többnyire csupán a szerkezet igeiségéért felel [1], mint például *irányt ad, forgalomba hoz* vagy *ajánlatot tesz*. Az FX-ek emellett szintaktikai, lexikai, szemantikai, pragmatikai vagy statisztikai szempontból idioszinkratikus tulajdonságokkal bírnak [2]. Ezen jellemzők, valamint relatív gyakoriságuk miatt tehát számos természetesnyelv-feldolgozó alkalmazás számára kulcsfontosságú lehet e szerkezetek folyó szövegben történő azonosítása.

A számítógépes nyelvészet számára ugyanakkor ez komoly kihívást jelent, hiszen az FX-ek (*segítséget kap*) felépítése szintaktikailag gyakorta egybeesik egyéb (kompozicionális) szerkezetekével (*pénzt kap*), valamint idiomatikus kifejezésekével (*vérszemet kap*). Másrészt mivel jelentésük nem teljes mértékben kompozicionális, így összetevőik különálló lefordítása is csak ritka esetben eredményezi az FX adott idegen nyelvi megfelelőjét. A nyelvek FX-einek eltérő sajátosságai pedig további nehézségeket jelenthetnek az automatikus azonosítás számára.

Ezen sajátosságok figyelembevételével törekedtünk jelen munkánkban az FX-ek négy nyelven történő automatikus azonosítására. Kiindulópontunkat a magyar, német, angol és spanyol jogi szövegekből felépülő 4FX elnevezésű párhuzamos korpusz képezte, amelyben első lépésként a félig kompozicionális szerkezetek kerültek manuális annotálásra. Ahogyan azt a következőkben bemutatjuk, a kézi annotáció által nem csak a különböző nyelvek FX-einek összehasonlítására nyílt lehetőségünk, hanem nyelvspecifikus tulajdonságaik feltárására is, melyek egy gépi tanuló alapjait képezték. Ezen adatvezérelt megközelítés a magyar nyelvre már bemutatott eljáráson alapszik [3], melyet sikeresen adaptáltunk a három másik nyelvre azok sajátosságainak figyelembevételével. A módszer szintaktikai elemzésre épülő FX-jelöltkiválasztó megközelítésre épül, mely a potenciális FX-ekről egy gazdag jellemzőterre támaszkodó gépi tanuló algoritmus segítségével hoz döntést.

## 2. Kapcsolódó munkák

A félig kompozicionális szerkezetek automatikus felismerésére, valamint a főnév + ige szerkezetek azonosítására már számos nyelvben kísérletet tettek, például az angolban [4–7], a hollandban [8], a németben [9], valamint a baszkban [10].

A többszavas kifejezések identifikálásában rendkívüli fontossággal bírnak a párhuzamos korpuszok. Ennek kapcsán Caseli és munkatársai [11] egy olyan osztályozáson alapuló módszert dolgoztak ki, mely portugál-angol párhuzamos korpuszból képes kinyerni az FX-eket.

Samardžić és Merlo [9] angol és német nyelvű párhuzamos szövegállományban található félig kompozicionális szerkezeteket vizsgálva jutott arra a megállapításra, hogy az FX-ek párhuzamosításánál a gyakorisági adatok mellett nyelvi jellemzők is fontos szerepet játszanak.

Zarriß és Kuhn [12] bemutatta, hogy a többszavas kifejezések hatékonyan detektálhatóak a parallel szövegekben fordítási párhuzamok alapján.

Attia és munkatársai [13] pedig arab többszavas kifejezések azonosításakor támaszkodtak a Wikipedia-bejegyzések párhuzamos címeiben található aszimmetriákra.

Ismereteink szerint az itt bemutatott az első olyan négynyelvű párhuzamos korpusz, amelyet a többszavas kifejezések egyidejű azonosítására használtak fel. A továbbiakban részletezzük a felhasznált korpusz tulajdonságait, valamint az FX-ek nyelvspecifikus sajátosságait.

## 3. A korpusz

A korpusz kialakítása során a JRC-Acquis [14] párhuzamos korpuszból indultunk ki, mely európai uniós jogi szövegeket tartalmaz. E szöveggyűjtemény angol nyelvű megfelelőjéből véletlenszerűen választottuk ki a szövegeket, amíg a korpusz mérete a százezer tokent nem haladta. Ezen angol nyelvű szövegek, valamint ezek német, magyar, illetve spanyol párhuzamos megfelelői kerültek

manuális annotálásra. Az így létrejövő korpusz képezte a manuális annotáció alapját. A műveletet két magyar anyanyelvű nyelvész végezte el, akik magas szintű német, angol és spanyol nyelvtudással rendelkeztek. Az egyes nyelveken annotált korpuszok méretét az 1. táblázat mutatja be.

1. táblázat. Az egyes részkorpuszok méretei.

	Angol	Német	Spanyol	Magyar	Összesen
Mondatok száma	5220	6392	5369	4927	21908
Szavak száma	100169	99258	111266	89338	400031

Ahogy a 1. táblázat mutatja, szavak számának tekintetében a német és az angol korpusz közel megegyező, a spanyol szövegállomány tokenszáma ennél csaknem 10 százalékkal több, míg a magyaré körülbelül ugyanennyivel kevesebb. A mondatok és szavak számát egybevetve ugyanakkor megfigyelhető, hogy az angolhoz képest a spanyol nyelvben jóval hosszabb mondatok jellemzőek, a németben inkább a „több rövidebb mondat” elve érvényesül, míg a magyar mondatok hosszúsága az angoléhoz közelít. A nyelvek közötti eltérések ugyanakkor nem csak e tekintetben válnak nyilvánvalóvá, hanem, amint azt a következőkben bemutatjuk, az FX-ek számát és felszíni formáját illetően is lényeges különbségek állapíthatók meg.

#### 4. Annotációs elvek

Az FX-ek minél egységesebb annotálása érdekében bizonyos alaptételeket tartottunk szem előtt. Ezek a SzegedParalellFX [15] kialakítása során alkalmazott irányelveket foglalták magukban, azaz olyan kérdéseket, mint például *A főnévi komponenssel morfológiailag megegyező tövű főige képes-e helyettesíteni a szerkezetet?*, *Az ige elhagyásával rekonstruálható-e az eredeti cselekvés?*, *A szerkezet nominalizálható, illetve passzivizálható-e?*. Ezen kérdéseket a magyar és az angol nyelv mellett a németben és a spanyolban is felhasználtuk.

A korpusz építése során a másik lényeges alapelv volt, hogy nem csupán a prototipikus felépítésű igei FX-eket jelöltük (VERB, pl. *forgalomba hoz*), hanem a melléknévi igenévi (PART, pl. *forgalomba hozott*), illetve a főnévi (NOM, pl. *forgalomba hozatal*) alakokat is. Emellett a félig kompozicionális szerkezetek nem folytonos változatait (SPLIT, pl. *hozta a vállalat forgalomba*) is bevontuk az annotálásba. Erre vonatkozó adatainkat a 2. táblázat mutatja be.

Az itt közölt gyakorisági statisztikák annál is inkább figyelemre méltóak, minthogy azonos szövegállomány különböző nyelvű párhuzamos variánsai képeztek kiindulópontunkat. Az adatokból kitűnik például, hogy a spanyol korpuszban csaknem kétszer annyi FX található, mint azok angol megfelelőiben. Ez pedig egyértelműen alátámasztja az FX-ek különbözőségét az annotálásba bevont nyelvek között, ugyanakkor a nyelvek között is különbségek tapasztalhatóak az FX-

2. táblázat. Manuálisan annotált FX-ek gyakoriságai különböző nyelveken.

	Angol	Német	Spanyol	Magyar	Összesen
NOM	24 5,47%	241 18,24%	73 8,34%	160 19,98%	498 17,24%
VERB	186 42,37%	216 27,94%	494 56,46%	300 37,45%	1196 41,42%
SPLIT	79 18,00%	214 27,68%	119 13,60%	68 8,49%	480 16,62%
PART	150 34,17%	102 13,20%	189 21,60%	273 34,08%	714 24,72%
Összesen	439 100,00%	773 100,00%	875 100,00%	801 100,00%	2888 100,00%

ek tekintetében. Ezen eltérések okainak és az FX-ek nyelvspecifikus jellemzőinek pontos feltárása pedig az első lépés lehet azok automatikus azonosításában.

## 5. FX-ek nyelvspecifikus sajátosságai

A kézi annotáció eredményeinek elemzése egyértelműen rámutat az imént említett nyelvspecifikus sajátosságokra.

A 2. táblázat egyik szembetűnő eredménye például, hogy a négy nyelv közül a németben a leggyakoribbak a nem folytonos FX-ek. Kötött szórendű nyelvről lévén szó itt az ige alapvetően a második helyen áll, argumentumainak pozíciója azonban már jóval rugalmasabb. Az FX-ek esetében ez azt eredményezi, hogy a főnévi komponens gyakran a mondat utolsó tagjaként mintegy keretes szerkezetet alkot az igével, pl.:

*Diese Verordnung tritt am 31. März 2006 in Kraft.*

*Ez a rendelet 2006. március 31-én lép hatályba.*

E tulajdonságának köszönhetően a németben a legmagasabb a SPLIT konstrukciók száma, melynek aránya megközelíti a folytonos szerkezetekét. Ugyanakkor a német nyelv további sajátossága, hogy a főnévi alakok (NOM) száma messze meghaladja a többi nyelvben találhatóét, mely jelenségre a német szakszövegekre gyakran jellemző nominális stílus (Nominalstil) adhat magyarázatot. Elemzéseink statisztikailag is alátámasztották tehát azt a tényt, amelyet a német szakirodalom az FX-ek kapcsán gyakorta megemlít: a jogi nyelvezet sajátja a főnevesítést előtérbe helyező kifejezésmód, melynek egyik legtipikusabb indikátora a félig kompozicionális szerkezetek alkalmazása is. Ezen értékeiben a magyar nyelvhez áll a legközelebb a német [16].

Ugyanakkor ezt leszámítva azonban nem állapítható meg nagy egyezés a magyar nyelvvel. Továbbá érdekes tény például, hogy a magyarban messze a legalacsonyabb a SPLIT-es szerkezetek aránya. Ennek oka lehet, hogy egyrészt

nincsen előre meghatározott szórend, és a szavak egymásutánisága a mondat információs struktúráját tükrözi, így a nem folytonos FX-ek esetében általában a közbeékelődő információra helyeződik a hangsúly. Valószínűleg a jogi szövegek tárgyilagosságra törekedve kerülhetik bizonyos információk kihangsúlyozását, melynek köszönhetően előnyben részesítik a folytonos FX-eket. A hangsúlyok eltolódását a következő mondatok jól szemléltetik:

*A bíróságot a kellő visszatartó hatásnak megfelelő mértékben szabják meg.*

*A kellő visszatartó hatásnak megfelelő mértékben szabják meg a bíróságot.*

Az adatok emellett szembetűnően mutatják, hogy a spanyol nyelv alkalmaz leggyakrabban félig kompozicionális szerkezeteket, melyek száma csaknem kétszerese az angol FX-ekének. A szerkezetek jelentős része folytonos, ennek kapcsán pedig egy különös sajátosságát is szükséges megemlíteni a spanyol FX-eknek. Korpuszunkban több példát is találtunk ugyanis a kettős FX-ekre, melyeket a következő szerkezetek példáznak:

*lleva a cabo la aproximación (közeledést hajt végre)  
da lugar a malentendidos (félreértéseknek ad helyt)*

Könnyen belátható, hogy a magyar nyelv számára sem idegen konstrukciókról van szó, mivel azonban ezekkel nem találkoztunk sem a német, sem az angol nyelvű korpusz annotálása során, így feltételezhető, hogy ténylegesen egy nyelvspecifikus tényezővel van dolgunk.

3. táblázat. Gépi tanuló megközelítés eredményei a különböző nyelveken.

	Szótárillesztés			Gépi tanuló		
	Pontosság	Fedés	F-mérték	Pontosság	Fedés	F-mérték
Angol	78,46	29,48	42,86	70,87	61,78	66,01
Német	82,5	7,61	13,92	58,81	46,91	52,19
Spanyol	57,22	32,71	41,65	65,7	45,48	53,75
Magyar	77,65	25,09	37,93	78,55	62,79	69,79

## 6. Gépi tanuló megközelítés az FX-ek automatikus azonosítására

Az FX-ek folyó szövegekben való automatikus azonosítására alapvetően a [3] megközelítést alkalmaztuk. A módszer először különböző morfológiai és szintaktikai jellemzőkre alapozó jelöltkiválasztó módszerek segítségével választja ki a potenciális FX-eket folyó szövegekből, majd egy gazdag jellemzőkészleten alapuló

döntési fa mesterséges intelligencia algoritmus alapján szelektálja ki a jelöltek közül az FX-eket. A módszert alapvetően angol, valamint magyar nyelvre valószínűsítették meg, továbbá az alap jellemzőkészlet mind a két nyelv esetében ki van egészítve nyelvspecifikus jellemzőkkel. Ezt a megközelítést alkalmaztuk az angol, valamint a magyar részkorpuszon, valamint adaptáltuk spanyol és német nyelvre. Ehhez jelöltkiválasztó módszereket definiáltunk a spanyol, valamint a német nyelvre, ami az angol és magyar nyelvű módszerekhez hasonlóan történt. Továbbá szükséges volt az alap jellemzőkészletet az aktuális nyelvhez igazítani és implementálni, valamint mind a két új nyelv esetében kiegészítettük a jellemzőkészletet nyelvspecifikus jellemzőkkel. Így a német és a spanyol esetében új morfológiai jellemzőként definiáltuk a főnevek nemét, míg német esetében az összetett főneveket. A rendszert minden nyelv esetében tízszeres keresztvalidációval értékeltük ki az aktuális részkorpuszon.

A gépi tanuló megközelítésünket minden nyelv esetében összevetettük egy szótárillesztési alpmegközelítéssel. Ebben az esetben azokat az FX-eket jelöltük, amelyeket a különböző jelöltkiválasztó algoritmusok választottak ki a folyószövegből, valamint egy adott FX listában szerepelnek. A megközelítés eredményei a 3. táblázatban találhatók.

## 7. Eredmények

Ahogy az a 3. táblázatban is látható, gépi tanuló megközelítésünk német és spanyol nyelven elért eredményei valamivel szerényebbek az angol és magyar nyelvű részkorpuszokon elért értékekhez képest. Ennek megfelelően a legjobb eredményeket a magyar és angol nyelvű részkorpuszon értünk el 69,79-os, valamint 66,01-os F-mértékkel, melyekhez viszonylag magas pontosságértékek tartoztak. Ezzel szemben német és spanyol nyelven csupán valamivel 50-es F-mértéket meghaladó eredményeket kaptunk, melyek elsősorban a gyenge fedési eredményeknek volt köszönhető. A német nyelvet leszámítva a szótáralapú megközelítés 40-es F-mérték körüli értékeket ért el. Továbbá érdemes megemlíteni, hogy a magyar és a spanyol nyelv esetében a gépi tanuló megközelítés magasabb pontosságértéket tudott elérni a szótárillesztésnél.

## 8. Az eredmények értékelése, összegzés

Jelen munkákban bemutatottuk 4FX elnevezésű korpuszunkat, melyben a JRC-Acquis párhuzamos, többnyelvű korpusz négy különböző nyelven manuálisan annotált FX-ei találhatók. A korpuszon egy már meglévő, gépi tanuló algoritmuson alapuló megközelítés segítségével automatikusan azonosítottuk a folyó szövegekben az FX-eket. Mivel a megközelítés korábban csak angol és magyar nyelvű FX-ek azonosítására volt képes, ezért szükséges volt azt spanyol és a német nyelvre adaptálni. Ahogy az a 3. táblázatban látható, az általunk alkalmazott gépi tanuló megközelítés robosztusnak tekinthető, mivel az négy különböző nyelven is képes volt felülmúlni a szótárillesztési alpmódszerünket. Ehhez a különböző nyelvspecifikus jellemzők is hozzájárultak.

Az egyes nyelvek közti egyértelmű eltérés alapvetően a fedésértékben mutatkozik meg. Így a német és spanyol nyelven elért gyengébb eredményekért elsősorban a fedésértékek felelnek, ami alapvetően a jellemzőkinyerő megközelítések gyengébb teljesítményének a következménye.

Német nyelvben az azonosításkor például komolyabb problémát jelentett a szabad szórend lehetőségéből fakadó nem folytonos szerkezetek magas száma, amit a szótárillesztő megközelítés meglehetősen alacsony fedésértéke is mutat. E szerény adatok azzal is magyarázhatóak továbbá, hogy bár a produktív módon képzett főnévi FX-ek a magyar mellett itt fordulnak elő a legnagyobb gyakorisággal, azonosításukra azonban még nincsen teljes mértékben felkészítve az itt bemutatott megközelítésünk.

A magyarban az azonosítási hibák a főnevek problematikája mellett főként a nyelv morfológiai sokszínűségéből fakadtak, hiszen itt az igei alakok a szám, személy, igeidő és igemód függvényében számos eltérő ragokat kaphatnak. Ugyanez érvényesnek tűnik a spanyol tekintetében is, ahol a morfológiai gazdagság miatt az igeiken túl a melléknévi igenevek azonosítása, valamint a korábban bemutatott kettős FX-ek felismerése is gyakori hibaforrásnak számít. Az angol nyelv esetében a hibák egy további jellemző csoportját szükséges kiemelniük, mégpedig a homonim alakokat. Itt ugyanis a szerkezet főnévi alakja (*to have a walk*) több esetben megegyezik a szerkezetet helyettesítő főigével (*to walk*), ami hibát jelenthet az automatikus szófaji egyértelműsítés számára, és ez szintén növeli a hibaforrások számát.

A nyelvek tehát szembetűnő eltéréseket mutatnak az FX-ek tekintetében, ami meglehetősen eltérő nyelvspecifikus jellemzők definiálását teszi szükségessé. Továbbá, ahogyan a 2. táblázat is mutatja, az FX-ek gyakorisága is jelentősen eltér a különböző nyelvekben. Összességében azonban megállapítható, hogy a nyelvi specifikumok ellenére is lehet létjogosultsága az általunk kidolgozott megközelítésnek, melynek további finomítása jövőbeli terveink között szerepel.

## Köszönetnyilvánítás

A kutatás a futurICT.hu nevű, TÁMOP-4.2.2.C-11/1/KONV-2012-0013 azonosítószerű projekt keretében zajlott. Nagy T. Istvánt a TÁMOP 4.2.4.A/2-11-1-2012-0001 azonosítószerű Nemzeti Kiválóság Program – Hazai hallgatói, illetve kutatói személyi támogatást biztosító rendszer kidolgozása és működtetése konvergencia program című kiemelt projekt támogatta. Mindkét projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

## Hivatkozások

1. Vincze, V.: Semi-Compositional Noun + Verb Constructions: Theoretical Questions and Computational Linguistic Analyses. PhD thesis, Szegedi Tudományegyetem, Szeged (2011)
2. Calzolari, N., Fillmore, C., Grishman, R., Ide, N., Lenci, A., MacLeod, C., Zampolli, A.: Towards best practice for multiword expressions in computational lexicons. In: Proceedings of LREC-2002, Las Palmas (2002) 1934–1940

3. Vincze, V., Nagy T., I., Farkas, R.: Identifying English and Hungarian Light Verb Constructions: A Contrastive Approach. In: *Proceedings of ACL (Volume 2: Short Papers)*, Sofia, Bulgaria, ACL (2013) 255–261
4. Cook, P., Fazly, A., Stevenson, S.: Pulling their weight: exploiting syntactic forms for the automatic identification of idiomatic expressions in context. In: *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions. MWE '07*, Morristown, NJ, USA, ACL (2007) 41–48
5. Bannard, C.: A measure of syntactic flexibility for automatically identifying multiword expressions in corpora. In: *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions. MWE '07*, Morristown, NJ, USA, ACL (2007) 1–8
6. Vincze, V., Nagy T., I., Berend, G.: Detecting Noun Compounds and Light Verb Constructions: a Contrastive Study. In: *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, Portland, Oregon, USA, ACL (2011) 116–121
7. Tu, Y., Roth, D.: Learning English Light Verb Constructions: Contextual or Statistical. In: *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, Portland, Oregon, USA, ACL (2011) 31–39
8. Van de Cruys, T., Moirón, B.V.: Semantics-based multiword expression extraction. In: *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions. MWE '07*, Morristown, NJ, USA, ACL (2007) 25–32
9. Samardžić, T., Merlo, P.: Cross-lingual variation of light verb constructions: Using parallel corpora and automatic alignment for linguistic research. In: *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, Uppsala, Sweden, ACL (2010) 52–60
10. Gurrutxaga, A., Alegria, I.: Automatic Extraction of NV Expressions in Basque: Basic Issues on Cooccurrence Techniques. In: *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World*, Portland, Oregon, USA, ACL (2011) 2–7
11. Caseli, H.d.M., Villavicencio, A., Machado, A., Finatto, M.J.: Statistically-driven alignment-based multiword expression identification for technical domains. In: *Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications*, Singapore, ACL (2009) 1–8
12. Zarrieß, S., Kuhn, J.: Exploiting Translational Correspondences for Pattern-Independent MWE Identification. In: *Proceedings of the Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications*, Singapore, ACL (2009) 23–30
13. Attia, M., Toral, A., Tounsi, L., Pecina, P., van Genabith, J.: Automatic Extraction of Arabic Multiword Expressions. In: *Proceedings of the 2010 Workshop on Multiword Expressions: from Theory to Applications*, Beijing, China, Coling 2010 Organizing Committee (2010) 19–27
14. Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufiş, D.: The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In: *Proceedings of LREC 2006*. (2006) 2142–2147
15. Vincze, V., Felvégi, Zs., R. Tóth, K.: Félig kompozicionális szerkezetek a Szeged-Paralell angol-magyar párhuzamos korpuszban. In Tanács, A., Vincze, V., eds.: *MSzNy 2010*, Szeged, Szegedi Tudományegyetem (2010) 91–101
16. Duden: *Der Duden in 12 Bänden. Das Standardwerk zur deutschen Sprache: Duden 06. Das Aussprachewörterbuch: Unerlässlich für die richtige Aussprache. Betonung ... Namen: Bd 6 (Duden Series Volume 6))*: Band 6. Gebundene Ausgabe (2006)